Mol Breeding (2009) 23:337–347
DOI 10.1007/s11032-008-9246-7

Mapping candidate genes for oleate biosynthesis and their association with unsaturated fatty acid seed content in soybean

Eleni Bachlava · Ralph E. Dewey · Joseph W. Burton · Andrea J. Cardinal

Received: 27 September 2008 / Accepted: 24 November 2008 / Published online: 7 December 2008

Abstract The development of high-oleate soybean germplasm is hindered by the lack of knowledge of the genetic factors controlling oleate phenotypic variation. In the present study, several candidate genes implicated in oleate biosynthesis were mapped and their cosegregation with oleate, linoleate and linolenate quantitative trait loci (QTLs) was investigated. *FAD2-2C*, a previously described ω-6 desaturase isoform, was localized on linkage group E; whereas, a novel *FAD2-2* isoform, designated as *FAD2-2D*, mapped on linkage group N. In addition, two isoforms were identified for the aminoalcoholphosphotransferase-encoding *GmAAPT1* gene, denoted *AAPT1a* and *AAPT1b*. A database query suggested that only one functional copy of the *FAD6* gene, encoding a plastid localized ω-6 desaturase, exists in the soybean genome. *AAPT1a* and *FAD6* mapped on linkage group D1b, 23.40 cM apart. Linolenate QTLs with minor effects were identified near the *FAD6* and *AAPT1a* markers in two segregating populations.

Keywords Soybean · Oleate biosynthesis · QTL mapping · Fatty acid content

Introduction

Soybean [Glycine max (L.) Merr.] producers and processors are currently concerned with the need to lower the levels of *trans* isomers of fatty acids in processed soybean oils in order to meet the guidelines of the United States Food and Drug Administration, retain consumer acceptance and enhance the competitive position of soybean oil in domestic and global markets (Wilson et al. 2002). *Trans* isomers are formed during hydrogenation, which confers long-term oxidative stability to soybean oil. The need for oil hydrogenation will be diminished if the levels of the polyunsaturated fatty acids, linoleate and linolenate, which are susceptible to autoxidation, are reduced. Concomitant elevation of oleate content can also contribute to the oxidative stability of soybean oil and the improvement of its nutritional value (Wilson 2004). However, the lack of knowledge of

Electronic supplementary material The online version of this article (doi:10.1007/s11032-008-9246-7) contains supplementary material, which is available to authorized users.

E. Bachlava · R. E. Dewey · J. W. Burton · A. J. Cardinal (✉)
Department of Crop Science, North Carolina State University, Campus Box 7620, Raleigh, NC 27695-7620, USA
e-mail: andrea_cardinal@ncsu.edu

Present Address:
E. Bachlava
Center for Applied Genetic Technologies, University of Georgia, 111 Riverbend Rd., Athens, GA 30602, USA
e-mail: bachlava@uga.edu

J. W. Burton
USDA-ARS, Soybean and Nitrogen Fixation Research Unit, 3127 Ligon St, Raleigh, NC 27607, USA

the genetic factors contributing to the phenotypic variation of oleate content hampers the development of high-oleate soybean lines with traditional breeding.

The seed-specific *FAD2-1* genes and the constitutively-expressed *FAD2-2* genes are primarily responsible for oleate biosynthesis in soybean. These *GmFAD2* genes (Fig. 1) encode oleoyl-PC ω-6 desaturase activity and catalyze the desaturation of oleate to linoleate in the endoplasmic reticulum (Heppard et al. 1996). *FAD2-1A*, *FAD2-1B* and *FAD2-2B* isoforms were previously studied for associations with oleate phenotypes in soybean populations segregating for oleate content (Bachlava et al. 2008). Unlike other oilseed crops (Hu et al. 2006; Schuppert et al. 2003), no major QTLs cosegregated with these three *GmFAD2* isoforms. A unique oleate QTL with moderate effects was detected in the proximity of the *FAD2-1B* isoform on linkage group I (Bachlava et al. 2008, 2009). Therefore, further research on other candidate genes implicated in oleate biosynthesis, such as the *FAD2-2C* isoform, and the *FAD6* and *AAPT1* genes, is critical in order to investigate their association with oleate content.

FAD2-2C has been previously identified, but was not mapped in the soybean genome (Bachlava et al. 2008; Schlueter et al. 2007). Interestingly, *FAD2-2C* was the only *GmFAD2* isoform that showed changes in transcript accumulation at different temperatures during seed development. The level of expression of *FAD2-2C* increased eightfold in developing pods grown at 18/12°C day/night temperature, in comparison to those grown at 32/28°C day/night temperatures (Schlueter et al. 2007). Due to the strong temperature effect on oleate content during the pod-filling stage, which coincides with the period of oil deposition, *FAD2-2C* is an ideal candidate to explain oleate phenotypic variation.

The desaturation of oleate to linoleate in the plastids is controlled by the *FAD6* gene (Fig. 1), encoding the plastidial ω-6 desaturase enzyme in soybean (Heppard et al. 1996; Hitz et al. 1994). Although fatty acids of seed triacylglycerols are primarily synthesized through the eukaryotic pathway in the endoplasmic reticulum, the prokaryotic pathway in the plastids also contributes to the polyunsaturated fatty acid biosynthesis (Browse and Somerville 1991). It has been suggested that the disruption of one of the pathways is compensated by fatty acid biosynthesis through the other pathway. Intermediates are transferred between the compartments, but enzymes reside exclusively in either the endoplasmic reticulum or the plastids (Somerville and Browse 1996). The expression pattern of the *FAD6* gene has not suggested changes in the transcript levels in response to different growth temperatures (Heppard et al. 1996).

In soybean, the *AAPT1* gene (Fig. 1) encodes an aminoalcoholphosphotransferase that catalyzes the conversion of diacylglycerol to phosphatidylcholine in the endoplasmic reticulum and has a putative role in the mobilization of fatty acids to phosphatidylcholine (Dewey et al. 1994). This enzymatic reaction is believed to be reversible, allowing fatty acids to reenter the diacylglycerol pools destined for triacylglycerol production (Slack et al. 1983, 1985). *AAPT1* gene activity has the potential to influence the fatty acid composition since the substrates of desaturase enzymes in the endoplasmic reticulum are fatty acids esterified to the first and second backbone positions of phosphatidylcholine (Ohlrogge and Browse 1995). Given the duplicated nature of the soybean genome (Shoemaker et al. 2006), *AAPT1* is probably represented by a gene family with more than one isoform. The unique *AAPT1* gene characterized previously was predominantly expressed both during seed development and vegetative growth and showed evidence of post-transcriptional regulation (Dewey et al. 1994).

This study utilized sequence information that recently became available from the Soybean Genome Project (Stacey et al. 2008) to: (1) determine the number of distinct isoforms of *FAD2-2*, *FAD6* and *AAPT1* genes; (2) map the putative isoforms of each gene in the soybean genome; and (3) investigate whether any isoform of the aforementioned candidate genes for oleate biosynthesis are associated with the mono- and polyunsaturated fatty acid phenotypes.

Materials and methods

Population development and experimental design

Two soybean populations, previously denoted as FAF and FAS (Bachlava et al. 2008), were developed by single seed descent (Brim 1966). The FAF population consisted of 118 F_5- derived lines from the cross

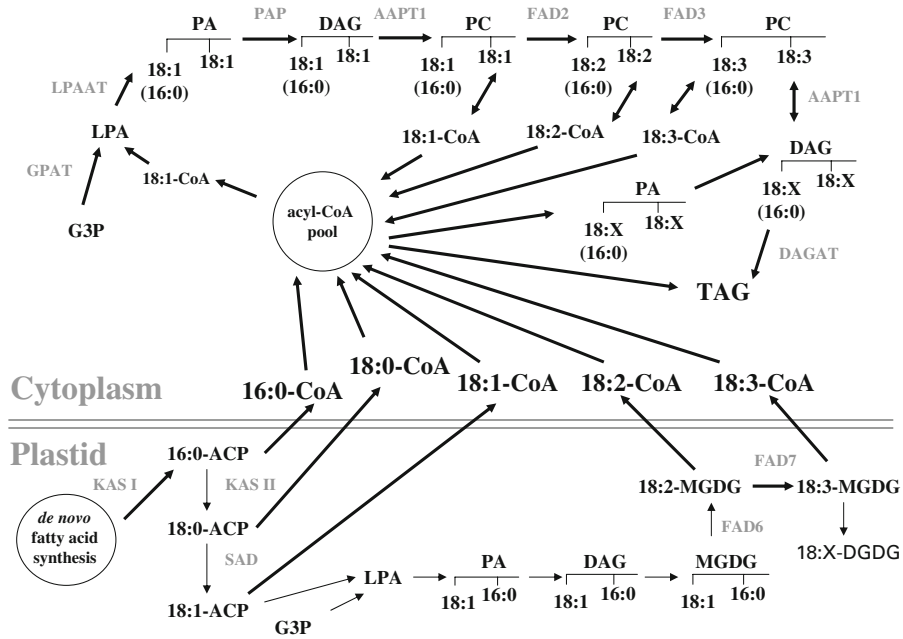

Fig. 1 Graphic representation of the fatty acid biosynthesis pathway, modified from Cardinal (2008). AAPT, aminoalcoholphosphotransferase; ACP, acyl carrier protein; DAG, diacylglycerol; DAGAT, diacylglycerol acyl transferase; DGDG, digalactosyl diacylglycerol; FAD, fatty acid desaturase; G3P, glycerol-3-phosphate; GPAT, glycerol-3-phosphate acyl transferase; KAS, 3-ketoacyl-ACP synthetase; LPA, lysophosphatidic acid; LPAAT, lysophosphatidic acid acyltransferase; MGDG, monogalactosyl diacylglycerol; PA, phosphatidic acid; PAP, phosphatidic acid phosphatase; PC, phosphatidylcholine; SAD, steroyl-ACP desaturase; TAG, triacylglycerol; 16:0, palmitate; 18:0, stearate; 18:1, oleate, 18:2, linoleate; 18:3, linolenate

of N97-3363-3 × PI423893 and the FAS population consisted of 231 F_3- derived lines from the cross of N98-4445A × PI423893. N98-4445A is a high-oleate (563.1 g kg^{-1}) low-linoleate line (Burton et al. 2006). N97-3363-3 is a sister line of N98-4445A and PI423893 is a mid-oleate (305.8 g kg^{-1}) plant introduction. Apart from oleate genes, the FAF and FAS populations were also segregating for the reduced linolenate *fan* (PI123440) allele.

The FAF population was planted in a sets within replications experimental design (Hallauer and Miranda 1988) with two replications in each location, as described by Bachlava et al. (2008). The FAF population was grown at Clinton and Kinston, NC, USA in 2005 and at Clinton, Kinston, Clayton and Plymouth, NC, USA in 2006. Each experimental line of the FAF population was planted in a four-row plot, with the exception of Clayton in 2006, where each line was planted in a one-row plot. The FAS population was planted in five-seed hills in a sets within replications design with three replications at Clayton in 2006 and 2007, as described by Bachlava et al. (2008).

Phenotypic evaluation and statistical analysis

Maturity date was recorded at the R8 reproductive stage (Fehr and Caviness 1977) as days after planting for all the environments where the FAF population was grown, and for the FAS population in 2006. Mature soybean seeds were harvested mechanically from the two middle rows of each experimental plot of the FAF population or from each one-row plot at Clayton in 2006. Approximately 10 g of seed were subsampled from each four-row plot for the evaluation of fatty acid composition. For the FAS population 5 g were subsampled from the seed harvested from each hill. Fatty acid composition was evaluated by gas liquid chromatography using a Model 6890 GC (Agilent Technologies, Inc., Wilmington, DE, USA), as described by Burkey et al. (2007).

Statistical analyses of the phenotypic data for the FAF and FAS populations were conducted using PROC MIXED in SAS 9.1 (SAS Institute Inc 2004). Environments, sets, replications, lines, and their interactions, were considered random effects. Best linear unbiased predictors (BLUPs) were obtained for maturity and fatty acid traits. For each experimental line of the FAF and FAS population, BLUPs were estimated as the sum of the intercept and the random effect (Littell et al. 1996). The phenotypic data of the FAF population from Clinton, NC, USA in 2005 were discarded due to excess missing data and greater error variance compared to the other environments. BLUPs were derived separately for each environment and combined across all environments for both the FAF and FAS populations.

'Gene-specific' SSR marker development and genotypic evaluation

Sequence information obtained from GenBank (Benson et al. 2002) accessions AC166742 and AB188252, which correspond to isoforms of the *GmFAD2-2* gene, and accessions L29215 and U12735, which correspond to the *GmFAD6* and *GmAAPT1* genes, were utilized for the BLAST application of the preliminary soybean assembly (Stacey et al. 2008). Scaffolds containing the *FAD2-2*, *FAD6* and *AAPT1* genes were identified. Sequence of ~100 Kb length, covering the region of the genome around each gene of interest, was mined for simple sequence repeats (SSRs) using SSRIT (Temnykh et al. 2001). Primers were designed to amplify the SSRs using Primer3 v. 0.4 (Rozen and Skaletsky 2000).

Polymorphic SSR markers were identified in the proximity of *FAD2-2*, *FAD6* and *AAPT1* genes (Table 1) and were used for genotyping of the FAF and FAS populations. The polymerase chain reactions (PCRs) were performed in a 384-well PTC-200 Peltier Thermal Cycler (MJ Research, Waltham, MA, USA). Amplification of genomic DNA was conducted with a 12-µl PCR reaction containing 6 pmol of each of the forward and reverse primers, 0.5 U Taq polymerase (New England Biolabs, Ipswich, MA, USA), 10× buffer (10 mM Tris–HCl, 50 mM KCl, 1.5 mM $MgCl_2$, pH 8.3), 234 µM dNTPs and 60 ng genomic DNA. Thermocycling conditions were 95°C for 2 min, 17 cycles of 92°C for 1 min, 67°C for 1 min, decreased by 1°C per cycle, and 68°C for 1 min 45 s, and 29 cycles of 92°C for 1 min, 49°C for 1 min and 68°C for 1 min 45 s. The amplification products were resolved on 4% SFR agarose gels (Amresco, Solon, OH, USA) with ethidium bromide staining in 1× Tris–Borate–EDTA buffer.

Amplification products of the polymorphic SSR markers using genomic DNA of the N98-4445A, N97-3363-3 and PI423893 lines were sequenced in order to verify their specificity. The PCR reactions were performed using the Expand High Fidelity PCR system (Roche Applied Science, Indianapolis, IN, USA) in a MyCycler thermal cycler (Bio-Rad, Hercules, CA, USA). Each 50 µl reaction contained 0.8 U polymerase, 20 pmol of each of the forward and reverse primers, 187 µM dNTPs and 100 ng of genomic template DNA, and thermocycling conditions were the same as above. The amplification products were cloned into pCR 2.1 cloning vectors (Invitrogen, Carlsbad, CA, USA) and retrieved sequences (Iowa State DNA facility, Ames, IA, USA) were compared to the GenBank accessions using pairwise alignments.

The FAF population was also genotyped with SSR markers covering the 20 linkage groups of soybean genome (Cregan et al. 1999; Song et al. 2004), as previously described by Bachlava et al. (2008); the FAS population was only genotyped with SSR markers in the proximity of candidate genes for oleate biosynthesis. A total of 165 and 100 polymorphic SSR markers were genotyped for the FAF and FAS populations, respectively.

Linkage mapping and QTL mapping

Linkage analysis of the FAF population was conducted with MAPMAKER/EXP 3.0 (Lander et al. 1987) according to Cardinal et al. (2001), using a minimum likelihood of odds (LOD) score of 3.0 and a maximum Kosambi distance of 40 cM. The F_5-derived lines were analyzed as recombinant inbred lines after discarding the heterozygote genotypes. Linkage analysis of the F_3-derived lines of FAS population was conducted with JoinMap 3.0 (Van Ooijen and Voorrips 2001). The linkage map was constructed using Kosambi's mapping function and linkage was declared with maximum recombination frequency of 0.4 and minimum LOD score of 3.0. The linkage map presented herein (Fig. 2)

Table 1 Polymorphic simple sequence repeat (SSR) markers used for mapping of the *FAD2-2*, *FAD6* and *AAPT1* genes in the soybean genome. The candidate gene (also used as name of the SSR marker), the di-nucleotide repeat, the distance of the repeat from the gene, the primers designed for polymerase chain reaction (PCR) amplification of the SSR and the approximate sizes of the amplified fragments are reported

Candidate gene	Repeat	SSR distance from gene (bp)	Forward SSR primer[a] (5′–3′)	Reverse SSR primer[a] (5′–3′)	PCR band size (bp)
FAD6	AC	17,183[b]	GCTCATTCACAGACATGGTTGC	GCATGACGGGCCTATATCAGAA	265
AAPT1a	AT	1,840[b]	GCTCATGATGAGAAAGTGTGGAAA	GCGCTGTCAGTCTTAGTTACG	240
AAPT1b	AT	5,341[c]	GCTGGACACTGGACAATGGACT	GCCTTTGAATGTCGGGATTCCT	360
FAD2-2C	AT	33,154[d]	CGGATAGAAATGGCATGGTA	GCTTAGCTCAAGAGAGTTGTCCTT	365
FAD2-2D	AT	42,587[b]	GCCAAGGTGCATACATACGACA	GCCAACAATTTCTTTCCCCTCT	220

[a] With the exception of the forward primer designed for *FAD2-2C*, all primers contain a GC clamp at the 5′ end

[b] The distance from the middle of the repeat, located upstream of the candidate gene, to the predicted translational start codon

[c] The distance from the middle of the repeat, located within the candidate gene, to the predicted translational start codon

[d] The distance from the middle of the repeat, located downstream of the candidate gene, to the predicted stop codon

corresponds to linkage analysis of the FAS population using JoinMap 3.0 software.

QTL analysis was conducted with WinQTL Cartographer version 2.5 (Wang et al. 2007) using BLUPs derived for each trait and experimental line of the FAF and FAS populations, both across environments and in each environment separately. The marker genotypes were initially tested for association with the phenotypic traits using single factor analysis (SFA), a simple linear regression approach (Table 2).

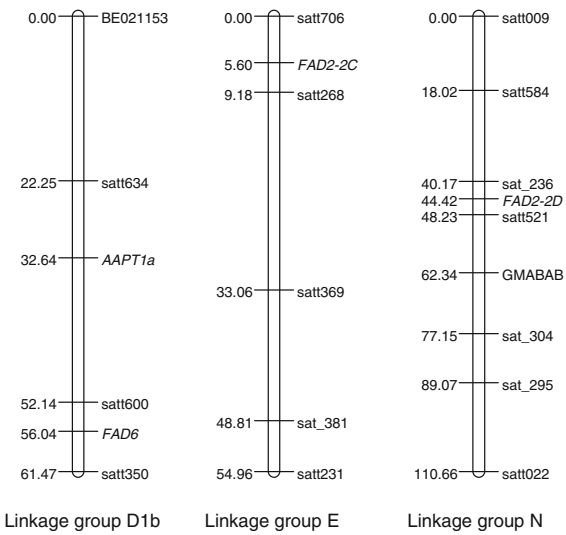

Fig. 2 Genetic map of soybean linkage groups D1b, E and N including the SSR markers developed for the *AAPT1a*, *FAD6*, *FAD2-2C* and *FAD2-2D* genes. Genetic distances were estimated with Kosambi's mapping function in the FAS population (N98-4445A × PI423893)

QTLs were identified in FAF and FAS populations using multiple interval mapping (MIM) (Kao et al. 1999). Initially, CIM procedure (Zeng 1993) was implemented with minimum LOD threshold of 2.5 to identify QTLs that are inputted in a starting model in MIM. The "forward and backward regression" method was used for cofactor selection and the genome scans were conducted with window size of 10 cM and walk speed of 2 cM. The CIM results for each trait and environment were inputted in the MIM procedure using the "scan through CIM analysis" option. The models were refined using the "optimize positions", "search for new QTLs" and "test existing QTLs" options, for main effects, in both populations, and epistatic interactions, in the FAS population. Due to the advanced inbreeding generation of the F_5-derived lines, only additive effects were fitted in the MIM models of the FAF population. MIM models with the minimum bayesian information criterion (BIC) were chosen and "search for new QTLs" was conducted until no additional QTLs could be detected. The QTL effects and the proportion of the variation explained from each QTL of the final MIM model was outputted using the "summary" option. Due to the increased statistical power of MIM for QTL mapping and the improved precision for estimation of QTL positions (Kao et al. 1999), only the results of MIM procedure were reported (Table 3). The QTLs presented in Table 3 were significant in MIM models that contained QTLs with moderate and minor effects mapping to several linkage groups of the soybean genome, which were

previously reported for the FAF and FAS populations (Bachlava et al. 2009), but are not presented in this study.

Results and discussion

Determining the isoforms of the *FAD2-2*, *FAD6* and *AAPT1* genes

Three isoforms of the *GmFAD2-2* gene have been previously identified. *FAD2-2B*, and the non-functional *FAD2-2A*, were localized on linkage group L, while *FAD2-2C* has not been mapped (Bachlava et al. 2008; Schlueter et al. 2007). Query of the sequence database of the Soybean Genome Project (Stacey et al. 2008) confirmed the existence of yet a fourth distinct isoform in the soybean genome, designated *FAD2-2D*, which corresponds to the GenBank accession AB188252. The identification of a new isoform was not surprising; Schlueter et al. (2007) predicted the existence of more than six *FAD2* loci in the soybean genome and Walling et al. (2006) found evidence of synteny for the long arm of linkage group L, where *FAD2-2A* and *FAD2-2B* mapped, with another chromosome. *FAD2-2D* displayed 96 and 95% similarity in the peptide and coding sequences, respectively, with the *FAD2-2B* isoform. Pertinent similarity evaluations for *FAD2-2C* with other *GmFAD2-2* isoforms was more problematic because the peptide predictions of the retrieved genome sequence suggested that *FAD2-2C* differed greatly from the *FAD2-2B* and *FAD2-2D* isoforms in the first 90 amino acids.

Pairwise alignment of coding, upstream and downstream regions suggested that *FAD2-2C* was likely derived from *FAD2-2B* or *FAD2-2D* through tandem duplication since it differed drastically in the upstream region and the initial 270 bp of the putative coding sequence, whereas the rest of the coding sequence (~880 bp) and part of the downstream region (~380 bp) were almost identical (Suppl. Fig. 1). Similarity of *FAD2-2B* and *FAD2-2D*, on the contrary, spanned the coding region and part of the downstream (~430 bp) and upstream (~1 Kb) regions of the isoforms (Suppl. Fig. 1). The fact that *FAD2-2B*, *FAD2-2C* and *FAD2-2D* have similar gene structure devoid of introns corroborates with the possibility of gene duplication. Schlueter et al. (2007) have previously discussed the possibility that *FAD2-2C* came from the tandem duplication of *FAD2-2B*, followed by genome rearrangements. Interestingly, we observed stop codons in all possible reading frames within this unique 5′ region of *FAD2-2C*, thus raising questions concerning the functionality of this gene. Although reverse transcriptase PCR experiments provided evidence that *FAD2-2C* is transcribed (Schlueter et al. 2007), we have not been able to identify any soybean expressed sequence tag (EST) that contains this putative *FAD2-2C*-specific sequence (data not shown).

Two isoforms were identified for the *AAPT1* gene within the soybean genome. Due to the temporary nomenclature of scaffolds in the preliminary soybean assembly (Stacey et al. 2008), the peptide, coding and full genome sequences are provided (Suppl. Fig. 2). The initial GenBank accession U12735, obtained by Dewey et al. (1994), corresponded to the *AAPT1b* isoform. Pairwise alignment revealed that the AAPT1a (389 amino acids) and AAPT1b (396 amino acids) predicted peptides share 97% identity. The yeast CPT1 peptide (GenBank accession

Table 2 Association (*P* values) of oleate, linoleate and linolenate content with markers developed for candidate genes of oleate biosynthesis tested in the FAF (N97-3363-3 × PI423893) and FAS (N98-4445A × PI423893) populations using single factor analysis (SFA)

FAF	Oleate	Linoleate	Linolenate	
	AAPT1a	*AAPT1a*	*AAPT1a*	*FAD6*
Across env[a]	0.049*		0.018*	0.013*
Kinston 2005	0.006**	0.011*	0.002**	0.008**
Kinston 2006			0.026*	0.010**
Clayton 2006			0.022*	0.017*
Clinton 2006				
Plymouth 2006				0.023*

FAS	Oleate		Linoleate	
	FAD2-2C	*AAPT1a*	*FAD2-2C*	*AAPT1a*
Across env	0.020*	0.011*	0.007**	0.012*
Clayton 2006		0.010**	0.019*	0.010**
Clayton 2007	0.015*	0.019*	0.006**	0.022*

*, ** Significance at the 0.05 and 0.01 level, respectively

[a] Oleate, linoleate and linolenate phenotypes used for SFA were derived as best linear unbiased predictors (BLUPs) both across environments and separately for each environment where the FAF and FAS populations were tested

Table 3 Quantitative trait loci (QTLs) for oleate, linoleate and linolenate detected in the proximity of candidate genes with multiple interval mapping (MIM) in the FAF (N97-3363-3 × PI423893) and FAS (N98-4445A × PI423893) populations. The linkage group (LG) and the marker interval where each QTL was mapped, as well as its size, the distance of the QTL peak from the most nearby marker (underlined), the additive (A) or dominance (D) effect of the QTL and the proportion of variation explained by each QTL (R^2) are presented. These findings are in addition to QTLs previously reported by Bachlava et al. (2009) for FAF and FAS populations in other linkage groups of soybean genome

Trait	Population	Environment[a]	LG	Marker interval	Interval size (cM)	Distance from nearby marker (cM)	Effect[b] (g kg^{-1})	R^2 (%)
Oleate	FAS	Across Env	D1b	*AAPT1a*—satt600	19.50	2.01	9.4 (A)	3.4
	FAS	Clayton 2006	D1b	satt634—*AAPT1a*	10.39	4.05	8.3 (A)	3.4
	FAF	Clinton 2006	E	*FAD2-2C*—satt706	3.70	0.01	13.5 (A)	4.5
Linoleate	FAS	Across Env	D1b	*AAPT1a*—satt600	19.50	1.54	−5.4 (A)	1.5
	FAS	Clayton 2006	D1b	satt634—*AAPT1a*	10.39	4.38	−6.9 (A)	3.6
	FAF	Clayton 2006	E	*FAD2-2C*—satt706	3.70	0.10	−8.9 (A)	1.5
	FAF	Clinton 2006	E	*FAD2-2C*—satt706	3.70	0.70	−9.2 (A)	3.6
Linolenate	FAS	Across Env	D1b	*FAD6*—satt350	5.43	0.01	5.7 (D)	4.6
	FAS	Clayton 2006	D1b	*FAD6*—satt350	5.43	1.86	7.3 (D)	6.2
	FAS	Clayton 2007	D1b	*FAD6*—satt350	5.43	1.16	5.4 (D)	4.3
	FAF	Across Env	D1b	*AAPT1a*—satt634	11.10	0.10	−1.8 (A)	3.7
	FAF	Kinston 2005	D1b	*AAPT1a*—satt634	11.10	0.10	−2.3 (A)	7.8
	FAF	Clayton 2006	D1b	*AAPT1a*—satt634	11.10	0.01	−2.2 (A)	3.8
	FAF	Plymouth 2006	D1b	*AAPT1a*—satt634	11.10	0.01	−1.4 (A)	2.2

[a] Oleate, linoleate and linolenate phenotypes used for MIM analysis were derived as best linear unbiased predictors (BLUPs) both across environments and separately for each environment where the FAF and FAS populations were tested

[b] Additive effect of the QTL was estimated in the FAF populations as the difference of the homozygous N97-3363-3 and PI423893 with MIM analysis. Positive additive effects in FAF indicated that the N97-3363-3 allele increases the value of the trait. Additive effect of the QTL was estimated in the FAS population as the difference of the homozygous N98-4445A and PI423893 genotypes and dominance effect was estimated as the difference of the heterozygous genotypes from the mean of the homozygous N98-4445A and PI423893 genotypes with MIM analysis. Positive additive effects in FAS indicated that the N98-4445A allele increases the value of the trait

NC_001146) was also aligned with the two AAPT1 peptides of soybean (Suppl. Fig. 3). A short-size polymorphism, seven amino acids in length, was found in the N-terminal regions of the AAPT1a and AAPT1b proteins. The fact that AAPT1a resembled the yeast CPT1 enzyme in this region suggests that an insertion event occurred in AAPT1b (Suppl. Fig. 3). Because of the high degree of sequence similarity, it is likely that both *AAPT1a* and *AAPT1b* encode functional isoforms of AAPT enzymes in soybean.

For the *FAD6* gene, a single functional isoform was detected in the initial draft of the soybean genome. The peptide and coding sequences were in agreement with the reports of Hitz et al. (1994), found in GenBank accession L29215. According to predictions of the preliminary soybean assembly (Stacey et al. 2008), the *FAD6* gene has nine introns with sizes ranging from 97 to 845 bp. A putative non-functional pseudogene with partial genomic sequence identity with the *FAD6* gene (~41% similarity of the predicted peptide sequences) was detected and mapped on linkage group D2, in the interval of the SSR markers satt328 and satt002 (data not shown).

Mapping the isoforms of *FAD2-2*, *FAD6* and *AAPT1* genes

The *FAD2-2C* and *FAD2-2D* isoforms mapped on linkage groups E and N of the soybean genome, respectively. *FAD2-2C* was localized in the interval of the SSR markers satt706 and satt268, 5.60 and 3.58 cM from the respective markers (Fig. 2). *FAD2-2D* mapped in the interval of sat_236 and satt521, 4.25 and 3.81 cM from each marker, respectively (Fig. 2). *FAD2-2C* and *FAD2-2D* did not map on the same linkage group, as may be predicted from the

extensive homeology of the long arm of linkage group L (chromosome 19), where *FAD2-2A* and *FAD2-2B* reside (Walling et al. 2006).

The *AAPT1a* isoform and the *FAD6* gene both mapped on linkage group D1b, 23.40 cM apart. *FAD6* was localized in the interval of satt600 and satt350, 3.90 and 5.43 cM from each marker, respectively (Fig. 2). The *AAPT1a* isoform mapped between satt634 and satt600, 10.39 cM and 19.50 cM from the respective markers (Fig. 2). The *AAPT1b* isoform could not be mapped, albeit the specific amplification of the *AAPT1b* marker was confirmed with sequencing. Linkage could not be declared with any of the 165 SSR markers genotyped for the FAF population or the 100 markers genotyped for the FAS population.

Associations of SSR markers with fatty acid content

According to the SFA analyses, the *AAPT1a* isoform and the *FAD6* gene on linkage group D1b were associated with changes in linolenate content in the FAF population. Both *AAPT1a* and *FAD6* markers were significant across environments and were respectively associated with linolenate content in three and four of the five environments of the FAF population (Table 2). In the FAS population, significant associations were identified for *FAD2-2C* and *AAPT1a* isoforms with oleate and linoleate content. Both SSR markers were significantly associated with linoleate content across environments and in each of the two environments where the FAS population was tested.

MIM analyses showed that the *AAPT1a* and *FAD6* genes on linkage group D1b cosegregated with QTLs controlling the unsaturated fatty acid content in soybean seed oil. A linolenate QTL was mapped near the *FAD6* marker in the FAS population. The QTL explained 4.6% of linolenate variation and the heterozygotes increased linolenate content by 5.7 g kg^{-1}, across environments, when compared to the mean of the two homozygotes for the N98-4445A and PI423893 alleles (Table 3). The QTL was present in each of the two environments of the FAS population and had similar dominance effects and R^2 values. Dominance effects were not fitted for the F_5- derived lines of the FAF population and, therefore, the linolenate QTL near *FAD6* could not be confirmed. This QTL was also identified using CIM analysis in the interval of the SSR markers *FAD6* and satt350 with a LOD score of 5.4, but due to the increased precision of QTL localization (Kao et al. 1999) only the MIM analysis is presented.

A linolenate QTL was also mapped to the *AAPT1a* marker in the FAF population. It explained 3.7% of linolenate variation across environments, and 2.2–7.8% of the observed variation in three of the five FAF environments (Table 3). The high-linolenate allele was inherited from PI423893 and led to an increase in linolenate content by 1.8 g kg^{-1}, across environments. The linolenate QTL was not confirmed in the FAS population; however, an oleate and linoleate QTL was identified near the *AAPT1a* marker both across environments and within one of the two environments, where the FAS population was tested. The direction of the additive effects of the oleate and linoleate QTL in the FAS population agreed with these of the linolenate QTL in the FAF population, since the high-oleate and high-linoleate alleles were inherited from N98-4445A and PI423893, respectively. The QTL near *AAPT1a* explained 3.4 and 1.5% of oleate and linoleate variation, respectively, and led accordingly to an increase in oleate content by 9.4 g kg^{-1} and in linoleate content by 5.4 g kg^{-1}, across environments (Table 3).

MIM analysis also revealed an oleate and linoleate QTL near *FAD2-2C* on linkage group E, which was present only in the 'Clinton 2006' environment of FAF population and explained 4.5 and 3.6% of oleate and linoleate genotypic variation (Table 3). Although this QTL was not confirmed with the MIM procedure for the FAS population, SFA suggested that the *FAD2-2C* marker is significantly associated with oleate and linoleate content in the FAS population. *FAD2-2C* mapped next to a genomic region where several QTLs for seed quality and agronomic traits are clustered. Minor QTLs have been previously reported on linkage group E for oleate and linolenate content near satt263 (Panthee et al. 2006), for linoleate content next to satt185 (Panthee et al. 2006), for maturity near satt263 (Panthee et al. 2006), and for seed weight near the SSR marker satt263 and the RFLP markers BLT049_5 and G214_26 (Fasoula et al. 2004; Mian et al. 1996; Orf et al. 1999). No QTL conditioning the unsaturated fatty acid content or maturity was reported in previous studies near the *FAD2-2D*, *AAPT1a* and *FAD6* genes on linkage groups N and D1b. In addition, no maturity QTLs

were identified in this study near the *AAPT1a*, *FAD6* and *FAD2-2C* markers; thus, there was no need to conduct additional QTL mapping analysis, as described by Bachlava et al. (2009), using BLUPs accounting for maturity effects.

Although the linolenate QTLs near *FAD6* and *AAPT1a* that were respectively identified in the FAS and FAF populations could not be confirmed, it is clear that minor QTLs implicated in the genetic control of the unsaturated fatty acid content in soybean oilseeds are present in the vicinity of the two candidate genes on linkage group D1b. The inconsistency in the QTLs identified for the FAF and FAS populations was probably due to their different sizes and levels of inbreeding, as well as the different environments where the FAF and FAS populations were grown (Beavis 1994; Bernardo 2002).

Conclusions

In the present study, and our previous work (Bachlava et al. 2008, 2009), we showed that naturally occurring alleles of several candidate genes implicated in oleate biosynthesis cosegregated with minor QTLs conditioning unsaturated fatty acid traits. Thereby, these alleles of candidate genes are probably responsible for only minor increases in the oleate content of the high-oleate lines N98-4445A and N97-3363-3. Our findings are in contrast to former studies on high-oleic lines carrying mutations in *FAD2-1* gene (Kinoshita et al. 1998; Sandhu et al. 2007) associated with major effects. Silencing of the *FAD2-1* gene also led to dramatic elevation in the levels of oleate content in the seed of transgenic soybean lines (Buhr et al. 2002; Kinney 1995; Kinney and Knowlton 1998). Yet, the elucidation of the genetic factors causing the observed natural variation in oleate content is necessary for understanding the biosynthesis of the high-oleate trait.

Acknowledgments The authors would like to thank William Novitzky at the USDA Soybean and Nitrogen Fixation Laboratory, Raleigh, NC, USA for providing training and equipment for the fatty acid analysis, as well as Carol Griffin for her technical expertise. This project was funded by the USDA Cooperative State Research, Education and Extension Service National Research Initiative, Plant Genome Program (award number 2003-35300-13190), and by the United Soybean Board.

References

Bachlava E, Dewey RE, Auclair J, Wang S, Burton JW, Cardinal AJ (2008) Mapping genes encoding microsomal omega-6 desaturase enzymes and their cosegregation with QTL affecting oleate content in soybean. Crop Sci 48:640–650. doi:10.2135/cropsci2007.07.0381

Bachlava E, Dewey RE, Burton JW, Cardinal AJ (2009) Mapping and comparison of quantitative trait loci for oleic acid seed content in two segregating soybean populations. Crop Sci (in press)

Beavis WD (1994) The power and deceit of QTL experiments: lessons from comparative QTL studies. In: Wilkinson DB (ed) 49th annual corn and sorghum research conference American seed trade association, Chicago

Benson DA, Karsch-Mizrachi I, Lipman DJ, Ostell J, Rapp BA, Wheeler DL (2002) GenBank. Nucleic Acids Res 30:17–20. doi:10.1093/nar/30.1.17

Bernardo R (2002) Breeding for quantitative traits in plants. Stemma Press, Woodbury

Brim CA (1966) A modified pedigree method of selection in soybean. Crop Sci 6:220

Browse J, Somerville C (1991) Glycerolipid synthesis: biochemistry and regulation. Annu Rev Plant Physiol Plant Mol Biol 42:467–506. doi:10.1146/annurev.pp.42.060191.002343

Buhr T, Sato S, Ebrahim F, Xing AQ, Zhou Y, Mathiesen M, Schweiger B, Kinney A, Staswick P, Clemente T (2002) Ribozyme termination of RNA transcripts down-regulate seed fatty acid genes in transgenic soybean. Plant J 30:155–163. doi:10.1046/j.1365-313X.2002.01283.x

Burkey KO, Booker FL, Pursley WA, Heagle AS (2007) Elevated carbon dioxide and ozone effects on peanut: II. Seed yield and quality. Crop Sci 47:1488–1497. doi:10.2135/cropsci2006.08.0538

Burton JW, Wilson RF, Rebetzke GJ, Pantalone VR (2006) Registration of N98-4445A mid-oleic soybean germplasm line. Crop Sci 46:1010–1012. doi:10.2135/cropsci2004-0769

Cardinal AJ (2008) Molecular genetics and breeding for fatty acid manipulation in soybean. Plant Breed Rev 30:259–294. doi:10.1002/9780470380130.ch5

Cardinal AJ, Lee M, Sharopova N, Woodman-Clikeman WL, Long MJ (2001) Genetic mapping and analysis of quantitative trait loci for resistance to stalk tunneling by the European corn borer in maize. Crop Sci 41:835–845

Cregan PB, Jarvik T, Bush AL, Shoemaker RC, Lark KG, Kahler AL, Kaya N, VanToai TT, Lohnes DG, Chung L, Specht JE (1999) An integrated genetic linkage map of the soybean genome. Crop Sci 39:1464–1490

Dewey RE, Wilson RF, Novitzky WP, Goode JH (1994) The *AAPT1* gene of soybean complements a cholinephosphotransferase-deficient mutant in yeast. Plant Cell 6:1495–1507

Fasoula VA, Harris DK, Boerma HR (2004) Validation and designation of quantitative trait loci for seed protein, seed oil, and seed weight from two soybean populations. Crop Sci 44:1218–1225

Fehr WR, Caviness CE (1977) Stages of soybean development. Spec Rep Agric Home Econ Experiment Stn Iowa State Univ 80:11

Hallauer AR, Miranda JB (1988) Quantitative genetics in maize breeding, 2nd edn. Iowa State University Press, Ames

Heppard EP, Kinney AJ, Stecca KL, Miao G (1996) Developmental and growth temperature regulation of two different microsomal omega-6 desaturase genes in soybeans. Plant Physiol 110:311–319. doi:10.1104/pp.110.1.311

Hitz WD, Carlson TJ, Booth R Jr, Kinney AJ, Stecca KL, Yadav NS (1994) Cloning of a higher-plant plastid omega-6 fatty acid desaturase cDNA and its expression in a cyanobacterium. Plant Physiol 105:635–641. doi:10.1104/pp.105.2.635

Hu X, Sullivan-Gilbert M, Gupta M, Thompson SA (2006) Mapping of the loci controlling oleic and linolenic acid contents and development of *fad2* and *fad3* allele-specific markers in canola (*Brassica napus* L.). Theor Appl Genet 113:497–507. doi:10.1007/s00122-006-0315-1

Kao CH, Zeng ZB, Teasdale RD (1999) Multiple interval mapping for quantitative trait loci. Genetics 152:1203–1216

Kinney AJ (1995) Improving soybean seed quality. Induced mutations and molecular techniques for crop improvement: In Proceedings. International Atomic Energy Agency (IAEA), Vienna, 19–23 June, pp 101-113

Kinney AJ, Knowlton S (1998) Designer oils: the high oleic acid soybean. In: Roller S, Harlander SK (eds) Genetic modification in the food industry. Chapman & Hall, London, pp 193–213

Kinoshita T, Rahman SM, Anai T, Takagi Y (1998) Genetic analysis of restriction fragment length polymorphism on fatty acid synthesis in soybean mutants and their progenies: II. High oleic acid mutants with two microsomal omega-6 fatty acid desaturase cDNAs as probes. Bull Fac Agric Saga Univ 83:37–42

Lander ES, Green P, Abrahamson J, Barlow A, Daly MJ, Lincoln ES, Newburg L (1987) MAPMAKER: an interactive computer package for constructing primary genetic linkage maps of experimental and natural populations. Genomics 1:174–181. doi:10.1016/0888-7543(87)90010-3

Littell RC, Milliken GA, Stroup WW, Wolfinger RD (1996) SAS® System for mixed models. SAS Institute Inc, Cary

Mian MAR, Bailey MA, Tamulonis JP, Shipe ER, Carter TE Jr, Parrott WA, Ashley DA, Hussey RS, Boerma HR (1996) Molecular markers associated with seed weight in two soybean populations. Theor Appl Genet 93:1011–1016. doi:10.1007/BF00230118

Ohlrogge J, Browse J (1995) Lipid biosynthesis. Plant Cell 7:957–970

Orf JH, Chase K, Jarvik T, Mansur LM, Cregan PB, Adler FR, Lark KG (1999) Genetics of soybean agronomic traits: I. Comparison of three related recombinant inbred populations. Crop Sci 39:1642–1651

Panthee DR, Pantalone VR, Saxton AM (2006) Modifier QTL for fatty acid composition in soybean oil. Euphytica 152:67–73. doi:10.1007/s10681-006-9179-3

Rozen S, Skaletsky HJ (2000) Primer3 on the www for general users and for biologist programmers. In: Krawetz S, Misener S (eds) Bioinformatics methods and protocols: methods in molecular biology. Humana Press, Totowa, pp 365–386

Sandhu D, Alt JL, Scherder CW, Fehr WR, Bhattacharyya MK (2007) Enhanced oleic acidcontent in the soybean mutant M23 is associated with the deletion in the Fad2-1a gene encoding a fatty acid desaturase. J Am Oil Chem Soc 84:229–235. doi:10.1007/s11746-007-1037-5

SAS Institute Inc (2004) SAS/STAT software. Release 9.1.3. SAS Institute Inc, Cary

Schlueter JA, Vasylenko-Sanders IF, Deshpande S, Yi J, Siegfried M, Roe BA, Schlueter SD, Scheffler BE, Shoemaker RC (2007) The *FAD2* gene family of soybean: insights into the structural and functional divergence of a paleopolyploid genome. Plant Genome, Suppl to Crop Sci 47:S14–S26

Schuppert GF, Heesacker A, Slabaugh MB, Cole G, Knapp SJ (2003) The high oleic acid phenotype in sunflower is caused by epistatic interactions between oleate desaturase genes. Plant & animal genome XI conference San Diego. http://wwwintl-pagorg/11/abstracts/W15_W101_XIhtml. Accessed 3 June 2008

Shoemaker RC, Schlueter J, Doyle JJ (2006) Paleopolyploidy and gene duplication in soybean and other legumes. Curr Opin Plant Biol 9:104–109. doi:10.1016/j.pbi.2006.01.007

Slack CR, Campbell LC, Browse JA, Roughan PG (1983) Some evidence for the reversibility of the cholinephosphotransferase-catalyzed reaction in developing linseed cotyledons in vivo. Biochim Biophys Acta 754:10–20

Slack CR, Roughan PG, Browse JA, Gardiner SE (1985) Some properties of cholinephosphotransferase from developing safflower cotyledons. Biochim Biophys Acta 833:438–448

Somerville C, Browse J (1996) Dissecting desaturation: plants prove advantageous. Trends Cell Biol 6:148–153. doi:10.1016/0962-8924(96)10002-7

Song QJ, Marek LF, Shoemaker RC, Lark KG, Concibido VC, Delannay X, Specht JE, Cregan PB (2004) A new integrated genetic linkage map of the soybean. Theor Appl Genet 109:122–128. doi:10.1007/s00122-004-1602-3

Stacey G, Shoemaker R, Jackson S, Schmutz J, Rokhsar D (2008) Soybean genome project, DoE Joint Genome Institute. http://wwwphytozomenet/soybeanphp. Accessed 3 June 2008

Temnykh S, DeClerck G, Lukashova A, Lipovich L, Cartinhour S, McCouch S (2001) Computational and experimental analysis of microsatellites in rice (*Oryza sativa* L.): frequency, length variation, transposon associations, and genetic marker potential. Genome Res 11:1441–1452. doi:10.1101/gr.184001

Van Ooijen JW, Voorrips RE (2001) Joinmap® 3.0: software for the calculation of genetic linkage maps. Plant Research International, Wageningen

Walling JG, Shoemaker R, Young N, Mudge J, Jackson S (2006) Chromosome-level homeology in paleopolyploid soybean (*Glycine max*) revealed through integration of genetic and chromosome maps. Genetics 172:1893–1900. doi:10.1534/genetics.105.051466

Wang S, Basten CJ, Zeng ZB (2007) Windows QTL Cartographer 2.5. Department of Statistics. North Carolina State University, Raleigh

Wilson RF (2004) Seed composition. In: Boerma HR, Specht JE (eds) Soybeans: improvement, production, and uses, 3rd edn. American Society of Agronomy Inc, Madison, pp 621–669

Wilson RF, Burton JW, Pantalone VR, Dewey RE (2002) New gene combinations governing saturated and unsaturated fatty acid composition in soybean. In: Kuo TM, Gardner HW (eds) Lipid Biotechnology. Marcel Dekker Inc, IL, pp 95–113

Zeng ZB (1993) Theoretical basis for separation of multiple linked gene effects in mapping quantitative trait loci. Proc Natl Acad Sci USA 90:10972–10976. doi:10.1073/pnas.90.23.10972